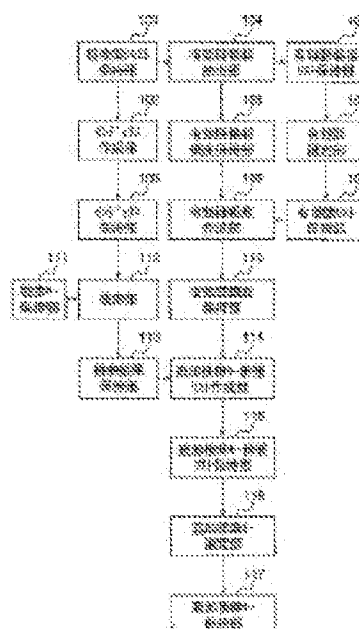


# **TEXT RETRIEVAL DEVICE, EFFECTIVE WORD FREQUENCY PREPARATION DEVICE, TEXT RETRIEVAL METHOD, EFFECTIVE WORD FREQUENCY PREPARATION METHOD AND RECORDING MEDIUM**

**Patent number:** JP2000112990 (A)  
**Publication date:** 2000-04-21  
**Inventor(s):** OTANI NORIKO; ITO SHIRO; UEDA TAKANARI; IKEDA YUJI +  
**Applicant(s):** CANON KK +  
**Classification:**  
- international: **G06F17/30; G06F17/30;** (IPC1-7): G06F17/30  
- european:  
**Application number:** JP19980300316 19981008  
**Priority number(s):** JP19980300316 19981008

## **Abstract of JP 2000112990 (A)**

**PROBLEM TO BE SOLVED:** To provide a text retrieval device and a text retrieval method which can easily re-retrieval matching a retrieval object document and a retrieval result document.  
**SOLUTION:** A block 104 extracts an effective word which becomes a candidate for an additional retrieval key from a retrieval object document, counts an appearance frequency in each document and prepares an effective word candidate list. Then, a block 107 refers to the effective word candidate list and prepares an effective word list. Furthermore, a block 109 deletes a word which is not an effective word from a document frequency of the effective word candidate held in a block 105 and prepares the document frequency of the effective word. A block 114 prepares an additional retrieval key candidate list from the one which corresponds to a document of the retrieval result out of the document frequency of the held effective word. A block 116 refers to the additional retrieval key candidate list and calculates an evaluation value for each additional retrieval key candidate. Then, the one whose-obtained evaluation value is high is presented as an additional retrieval key.



Data supplied from the **espacenet** database — Worldwide

(11)特許出願公開番号  
特開2000-112990  
(P2000-112990A)

310A 5B075  
340E

9/8/2009, EAST Version: 2.4.1.1

## 【特許請求の範囲】

【請求項1】 被検索テキストに対してインデックスを作成するインデックス作成手段と、

前記被検索テキストから有効語候補を抽出し、各文書における出現頻度をカウントして有効語候補リストを作成する有効語候補抽出手段と、

前記有効語候補抽出手段により抽出された有効語候補の出現頻度を文書別に保持する有効語候補頻度保持手段と、

前記有効語候補リストを参照して、各有効語候補の評価値を算出し、評価値の高い有効語候補を有効語として、有効語リストを作成する有効語選定手段と、

前記有効語リストを参照し、前記有効語候補頻度保持手段に保持された有効語候補の文書別頻度から有効語でない語を削除して、有効語の文書別頻度を作成する有効語頻度作成手段と、

前記インデックスにより、検索文字列に一致する文字列を検索する検索手段と、

前記有効語頻度保持手段に保持された有効語の文書別頻度のうち、前記検索手段の検索結果の文書に相当するものから、追加検索キー候補リストを作成する追加検索キー候補リスト作成手段と、

前記追加検索キー候補リストを参照して、各追加検索キー候補の評価値を算出する追加検索キー選定手段と、前記追加検索キー選定手段により求めた評価値が高いものを追加検索キーとして出力する追加検索キー提示手段とを備えたことを特徴とするテキスト検索装置。

【請求項2】 被検索テキストを保持する被検索テキスト保持手段と、

前記被検索テキスト保持手段に保持されている被検索テキストに対して、インデックスを作成するインデックス作成手段と、

前記インデックス作成手段で作成したインデックスを保持するインデックス保持手段と、

前記被検索テキスト保持手段に保持されている被検索テキストから有効語候補を抽出し、各文書における出現頻度をカウントして有効語候補リストを作成する有効語候補抽出手段と、

前記有効語候補抽出手段により抽出された有効語候補の出現頻度を文書別に保持する有効語候補頻度保持手段と、

前記有効語候補抽出手段により作成された有効語候補リストを保持する有効語候補リスト保持手段と、

前記有効語候補リスト保持手段に保持された有効語候補リストを参照して、各有効語候補の評価値を算出し、評価値の高い有効語候補を有効語として有効語リストを作成する有効語選定手段と、

前記有効語選定手段により作成された有効語リストを保持する有効語リスト保持手段と、

前記有効語リスト保持手段に保持された有効語リストを

参照し、上記有効語候補頻度保持手段に保持された有効語候補の文書別頻度から有効語でない語を削除して、有効語の文書別頻度を作成する有効語頻度作成手段と、

前記有効語頻度作成手段により作成された有効語の文書別頻度を保持する有効語頻度保持手段と、

検索を行なう文字列を保持する検索キー保持手段と、

前記インデックス保持手段に保持されているインデックスを用いて、前記検索キー保持手段に保持されている検索文字列に一致する文字列を検索する検索手段と、

前記検索手段により検索された検索結果を保持する検索結果保持手段と、

前記有効語頻度保持手段に保持された有効語の文書別頻度のうち、前記検索結果保持手段に保持された検索結果の文書に相当するものから、追加検索キー候補リストを作成する追加検索キー候補リスト作成手段と、

前記追加検索キー候補リスト作成手段により作成された追加検索キー候補リストを保持する追加検索キー候補リスト保持手段と、

前記追加検索キー候補リスト保持手段に保持された追加検索キー候補リストを参照して、各追加検索キー候補の評価値を算出する追加検索キー選定手段と、

前記追加検索キー選定手段により求めた評価値が高いものを追加検索キーとして保持する追加検索キー保持手段と、

追加検索キー保持手段に保持された追加検索キーを出力する追加検索キー提示手段とを備えたことを特徴とするテキスト検索装置。

【請求項3】 テキストから有効語候補を抽出し、各文書における出現頻度をカウントして有効語候補リストを作成する有効語候補抽出手段と、

前記有効語候補抽出手段により抽出された有効語候補の出現頻度を文書別に保持する有効語候補頻度保持手段と、

前記有効語候補リストを参照して各有効語候補の評価値を算出し、評価値の高い有効語候補を有効語として有効語リストを作成する有効語選定手段と、

前記有効語リストを参照し、前記有効語候補頻度保持手段に保持された有効語候補の文書別頻度から有効語でない語を削除して、有効語の文書別頻度を作成する有効語頻度作成手段とを備えたことを特徴とする有効語頻度作成装置。

【請求項4】 被検索テキストに対してインデックスを作成するインデックス作成手段と、

前記インデックスにより、検索文字列に一致する文字列を検索する検索手段と、

請求項3記載の有効語頻度作成装置によって作成された有効語の文書別頻度のうち、前記検索手段の検索結果の文書に相当するものから、追加検索キー候補リストを作成する追加検索キー候補リスト作成手段と、

前記追加検索キー候補リストを参照して、各追加検索キー候補の評価値を算出する追加検索キー選定手段と、

前記追加検索キー選定手段により求めた評価値が高いものを追加検索キーとして出力する追加検索キー提示手段とを備えたことを特徴とするテキスト検索装置。

【請求項5】 被検索テキストに対してインデックスを作成するインデックス作成行程と、

前記被検索テキストから有効語候補を抽出し、各文書における出現頻度をカウントして有効語候補リストを作成する有効語候補抽出行程と、

前記有効語候補抽出行程により抽出された有効語候補の出現頻度を文書別に保持する有効語候補頻度保持行程と、

前記有効語候補リストを参照して、各有効語候補の評価値を算出し、評価値の高い有効語候補を有効語として、有効語リストを作成する有効語選定行程と、

前記有効語リストを参照し、前記有効語候補頻度保持行程で保持された有効語候補の文書別頻度から有効語でない語を削除して、有効語の文書別頻度を作成する有効語頻度作成行程と、

前記インデックスにより、検索文字列に一致する文字列を検索する検索行程と、

前記有効語頻度保持行程で保持された有効語の文書別頻度のうち、前記検索行程の検索結果の文書に相当するものから、追加検索キー候補リストを作成する追加検索キー候補リスト作成行程と、

前記追加検索キー候補リストを参照して、各追加検索キー候補の評価値を算出する追加検索キー選定行程と、

前記追加検索キー選定行程により求めた評価値が高いものを追加検索キーとして出力する追加検索キー提示行程とを実行することを特徴とするテキスト検索方法。

【請求項6】 被検索テキストを保持する被検索テキスト保持行程と、

前記被検索テキスト保持行程で保持されている被検索テキストに対して、インデックスを作成するインデックス作成行程と、

前記インデックス作成行程で作成したインデックスを保持するインデックス保持行程と、

前記被検索テキスト保持行程で保持されている被検索テキストから有効語候補を抽出し、各文書における出現頻度をカウントして有効語候補リストを作成する有効語候補抽出行程と、

前記有効語候補抽出行程により抽出された有効語候補の出現頻度を文書別に保持する有効語候補頻度保持行程と、

前記有効語候補抽出行程により作成された有効語候補リストを保持する有効語候補リスト保持行程と、

前記有効語候補リスト保持行程で保持された有効語候補リストを参照して、各有効語候補の評価値を算出し、評価値の高い有効語候補を有効語として有効語リストを作成する有効語選定行程と、

前記有効語選定行程により作成された有効語リストを保

持する有効語リスト保持行程と、

前記有効語リスト保持行程で保持された有効語リストを参照し、上記有効語候補頻度保持行程で保持された有効語候補の文書別頻度から有効語でない語を削除して、有効語の文書別頻度を作成する有効語頻度作成行程と、

前記有効語頻度作成行程により作成された有効語の文書別頻度を保持する有効語頻度保持行程と、

検索を行なう文字列を保持する検索キー保持行程と、

前記インデックス保持行程で保持されているインデックスを用いて、前記検索キー保持行程で保持されている検索文字列に一致する文字列を検索する検索行程と、

前記検索行程により検索された検索結果を保持する検索結果保持行程と、

前記有効語頻度保持行程で保持された有効語の文書別頻度のうち、前記検索結果保持行程で保持された検索結果の文書に相当するものから、追加検索キー候補リストを作成する追加検索キー候補リスト作成行程と、

前記追加検索キー候補リスト作成行程により作成された追加検索キー候補リストを保持する追加検索キー候補リスト保持行程と、

前記追加検索キー候補リスト保持行程で保持された追加検索キー候補リストを参照して、各追加検索キー候補の評価値を算出する追加検索キー選定行程と、

前記追加検索キー選定行程により求めた評価値が高いものを追加検索キーとして保持する追加検索キー保持行程と、

追加検索キー保持行程で保持された追加検索キーを出力する追加検索キー提示行程とを実行することを特徴とするテキスト検索方法。

【請求項7】 テキストから有効語候補を抽出し、各文書における出現頻度をカウントして有効語候補リストを作成する有効語候補抽出行程と、

前記有効候補抽出行程により抽出された有効語候補の出現頻度を文書別に保持する有効語候補頻度保持行程と、

前記有効語候補リストを参照して、各有効語候補の評価値を算出し、評価値の高い有効語候補を有効語として、有効語リストを作成する有効語選定行程と、

前記有効語リストを参照し、前記有効語候補頻度保持行程で保持された有効語候補の文書別頻度から有効語でない語を削除して、有効語の文書別頻度を作成する有効語頻度作成行程とを実行することを特徴とする有効語頻度作成方法。

【請求項8】 被検索テキストに対してインデックスを作成するインデックス作成行程と、

前記インデックスにより、検索文字列に一致する文字列を検索する検索行程と、

請求項7記載の有効語頻度作成方法によって作成された有効語の文書別頻度のうち、前記検索行程の検索結果の文書に相当するものから、追加検索キー候補リストを作成する追加検索キー候補リスト作成行程と、

前記追加検索キー候補リストを参照して、各追加検索キー候補の評価値を算出する追加検索キー選定行程と、前記追加検索キー選定行程により求めた評価値が高いものを追加検索キーとして出力する追加検索キー提示行程とを実行することとを特徴とするテキスト検索方法。

【請求項9】 被検索テキストに対してインデックスを作成するインデックス作成行程と、

前記被検索テキストから有効語候補を抽出し、各文書における出現頻度をカウントして有効語候補リストを作成する有効語候補抽出行程と、

前記有効語候補抽出行程により抽出された有効語候補の出現頻度を文書別に保持する有効語候補頻度保持行程と、

前記有効語候補リストを参照して、各有効語候補の評価値を算出し、評価値の高い有効語候補を有効語として、有効語リストを作成する有効語選定行程と、

前記有効語リストを参照し、前記有効語候補頻度保持行程で保持された有効語候補の文書別頻度から有効語でない語を削除して、有効語の文書別頻度を作成する有効語頻度作成行程と、

前記インデックスにより、検索文字列に一致する文字列を検索する検索行程と、

前記有効語頻度保持行程で保持された有効語の文書別頻度のうち、前記検索行程の検索結果の文書に相当するものから、追加検索キー候補リストを作成する追加検索キー候補リスト作成行程と、

前記追加検索キー候補リストを参照して、各追加検索キー候補の評価値を算出する追加検索キー選定行程と、

前記追加検索キー選定行程により求めた評価値が高いものを追加検索キーとして出力する追加検索キー提示行程とを含むプログラムを格納したことを特徴とするコンピュータ読み取り可能な記録媒体。

【請求項10】 被検索テキストを保持する被検索テキスト保持行程と、

前記被検索テキスト保持行程で保持されている被検索テキストに対して、インデックスを作成するインデックス作成行程と、

前記インデックス作成行程で作成したインデックスを保持するインデックス保持行程と、

前記被検索テキスト保持行程で保持されている被検索テキストから有効語候補を抽出し、各文書における出現頻度をカウントして有効語候補リストを作成する有効語候補抽出行程と、

前記有効語候補抽出行程により抽出された有効語候補の出現頻度を文書別に保持する有効語候補頻度保持行程と、

前記有効語候補抽出行程により作成された有効語候補リストを保持する有効語候補リスト保持行程と、

前記有効語候補リスト保持行程で保持された有効語候補リストを参照して、各有効語候補の評価値を算出し、評

10

20

30

40

50

価値の高い有効語候補を有効語として有効語リストを作成する有効語選定行程と、

前記有効語選定行程により作成された有効語リストを保持する有効語リスト保持行程と、

前記有効語リスト保持行程で保持された有効語リストを参照し、上記有効語候補頻度保持行程で保持された有効語候補の文書別頻度から有効語でない語を削除して、有効語の文書別頻度を作成する有効語頻度作成行程と、

前記有効語頻度作成行程により作成された有効語の文書別頻度を保持する有効語頻度保持行程と、

検索を行なう文字列を保持する検索キー保持行程と、

前記インデックス保持行程で保持されているインデックスを用いて、前記検索キー保持行程で保持されている検索文字列に一致する文字列を検索する検索行程と、

前記検索行程により検索された検索結果を保持する検索結果保持行程と、

前記有効語頻度保持行程で保持された有効語の文書別頻度のうち、前記検索結果保持行程で保持された検索結果の文書に相当するものから、追加検索キー候補リストを作成する追加検索キー候補リスト作成行程と、

前記追加検索キー候補リスト作成行程により作成された追加検索キー候補リストを保持する追加検索キー候補リスト保持行程と、

前記追加検索キー候補リスト保持行程で保持された追加検索キー候補リストを参照して、各追加検索キー候補の評価値を算出する追加検索キー選定行程と、

前記追加検索キー選定行程により求めた評価値が高いものを追加検索キーとして保持する追加検索キー保持行程と、

追加検索キー保持行程で保持された追加検索キーを出力する追加検索キー提示行程とを含むプログラムを格納したことを特徴とするコンピュータ読み取り可能な記録媒体。

【請求項11】 テキストから有効語候補を抽出し、各文書における出現頻度をカウントして有効語候補リストを作成する有効語候補抽出行程と、

前記有効語候補抽出行程により抽出された有効語候補の出現頻度を文書別に保持する有効語候補頻度保持行程と、

前記有効語候補リストを参照して、各有効語候補の評価値を算出し、評価値の高い有効語候補を有効語として、有効語リストを作成する有効語選定行程と、

前記有効語リストを参照し、前記有効語候補頻度保持行程で保持された有効語候補の文書別頻度から有効語でない語を削除して、有効語の文書別頻度を作成する有効語頻度作成行程とを含むプログラムを格納したことを特徴とするコンピュータ読み取り可能な記録媒体。

【請求項12】 被検索テキストに対してインデックスを作成するインデックス作成行程と、

前記インデックスにより、検索文字列に一致する文字列を検索する検索行程と、

請求項1.1記載の記録媒体中の有効語頻度作成行程によって作成された有効語の文書別頻度のうち、前記検索行程の検索結果の文書に相当するものから、追加検索キー候補リストを作成する追加検索キー候補リスト作成行程と、前記追加検索キー候補リストを参照して、各追加検索キー候補の評価値を算出する追加検索キー選定行程と、前記追加検索キー選定行程により求めた評価値が高いものを追加検索キーとして出力する追加検索キー提示行程とを含むプログラムを格納したことを特徴とするコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、検索対象文書から検索キーに対応した文書を提示するテキスト検索装置及びテキスト検索方法と、テキストにおける有効語の文書別頻度を作成する有効語頻度作成装置及び有効語頻度作成方法並びに記録媒体に関する。

【0002】

【従来の技術】一般に、テキスト検索装置では、検索対象文書から与えられた検索キーを含む文書を検索結果として提示する。

【0003】

【発明が解決しようとする課題】しかしながら、上記従来の装置では、大量に得られた検索結果を絞り込む場合や、関連する情報をさらに集めたい場合には、適切な検索キーを追加して検索をやり直さなくてはならないという問題があった。

【0004】そこで、検索キーに関連する語を追加検索キーとして提示し、再検索を容易にすることが考えられる。しかし、あらかじめ用意された辞書から検索キーの関連語を選出するのは、辞書にない検索キーに対応できなかったり、関連語を追加しても結果があまり変わらなかったりする。また、関連語は検索対象文書によっても異なり、検索対象文書に合わせて追加検索キーの候補を決める必要がある。すなわち、再検索は、検索対象文書と検索結果文書に適合して行う必要があるが、従来ではこれを容易に行う手法がなかった。

【0005】本発明は上記従来の問題点に鑑み、検索対象文書と検索結果文書に適合した再検索が容易に行なえるテキスト検索装置及びテキスト検索方法を提供することを目的とする。また、テキストにおける有効語の文書別頻度を簡易かつ的確に作成する有効語頻度作成装置及び有効語頻度作成方法並びに記録媒体を提供することを目的とする。

【0006】

【課題を解決するための手段】上記目的を達成するために第1の発明は、被検索テキストに対してインデックスを作成するインデックス作成手段と、前記被検索テキストから有効語候補を抽出し、各文書における出現頻度を

カウントして有効語候補リストを作成する有効語候補抽出手段と、前記有効語候補抽出手段により抽出された有効語候補の出現頻度を文書別に保持する有効語候補頻度保持手段と、前記有効語候補リストを参照して、各有効語候補の評価値を算出し、評価値の高い有効語候補を有効語として、有効語リストを作成する有効語選定手段と、前記有効語リストを参照し、前記有効語候補頻度保持手段に保持された有効語候補の文書別頻度から有効語でない語を削除して、有効語の文書別頻度を作成する有効語頻度作成手段と、前記インデックスにより、検索文字列に一致する文字列を検索する検索手段と、前記有効語頻度保持手段に保持された有効語の文書別頻度のうち、前記検索手段の検索結果の文書に相当するものから、追加検索キー候補リストを作成する追加検索キー候補リスト作成手段と、前記追加検索キー候補リストを参照して、各追加検索キー候補の評価値を算出する追加検索キー選定手段と、前記追加検索キー選定手段により求めた評価値が高いものを追加検索キーとして出力する追加検索キー提示手段とを備えたものである。

【0007】第2の発明では、被検索テキストを保持する被検索テキスト保持手段と、前記被検索テキスト保持手段に保持されている被検索テキストに対して、インデックスを作成するインデックス作成手段と、前記インデックス作成手段で作成したインデックスを保持するインデックス保持手段と、前記被検索テキスト保持手段に保持されている被検索テキストから有効語候補を抽出し、各文書における出現頻度をカウントして有効語候補リストを作成する有効語候補抽出手段と、前記有効語候補抽出手段により抽出された有効語候補の出現頻度を文書別に保持する有効語候補頻度保持手段と、前記有効語候補抽出手段により作成された有効語候補リストを保持する有効語候補リスト保持手段と、前記有効語候補リスト保持手段に保持された有効語候補リストを参照して、各有効語候補の評価値を算出し、評価値の高い有効語候補を有効語として有効語リストを作成する有効語選定手段と、前記有効語選定手段により作成された有効語リストを保持する有効語リスト保持手段と、前記有効語リスト保持手段に保持された有効語リストを参照し、上記有効語候補頻度保持手段に保持された有効語候補の文書別頻度から有効語でない語を削除して、有効語の文書別頻度を作成する有効語頻度作成手段と、前記有効語頻度作成手段により作成された有効語の文書別頻度を保持する有効語頻度保持手段と、検索を行なう文字列を保持する検索キー保持手段と、前記インデックス保持手段に保持されているインデックスを用いて、前記検索キー保持手段に保持されている検索文字列に一致する文字列を検索する検索手段と、前記検索手段により検索された検索結果を保持する検索結果保持手段と、前記有効語頻度保持手段に保持された有効語の文書別頻度のうち、前記検索結果保持手段に保持された検索結果の文書に相当するもの

から、追加検索キー候補リストを作成する追加検索キー候補リスト作成手段と、前記追加検索キー候補リスト作成手段により作成された追加検索キー候補リストを保持する追加検索キー候補リスト保持手段と、前記追加検索キー候補リスト保持手段に保持された追加検索キー候補リストを参照して、各追加検索キー候補の評価値を算出する追加検索キー選定手段と、前記追加検索キー選定手段により求めた評価値が高いものを追加検索キーとして保持する追加検索キー保持手段と、追加検索キー保持手段に保持された追加検索キーを出力する追加検索キー提示手段とを備えたものである。

【0008】第3の発明では、テキストから有効語候補を抽出し、各文書における出現頻度をカウントして有効語候補リストを作成する有効語候補抽出手段と、前記有効語候補抽出手段により抽出された有効語候補の出現頻度を文書別に保持する有効語候補頻度保持手段と、前記有効語候補リストを参照して各有効語候補の評価値を算出し、評価値の高い有効語候補を有効語として有効語リストを作成する有効語選定手段と、前記有効語リストを参照し、前記有効語候補頻度保持手段に保持された有効語候補の文書別頻度から有効語でない語を削除して、有効語の文書別頻度を作成する有効語頻度作成手段とを備えたものである。

【0009】第4の発明では、被検索テキストに対してインデックスを作成するインデックス作成手段と、前記インデックスにより、検索文字列に一致する文字列を検索する検索手段と、第3の発明の有効語頻度作成装置によって作成された有効語の文書別頻度のうち、前記検索手段の検索結果の文書に相当するものから、追加検索キー候補リストを作成する追加検索キー候補リスト作成手段と、前記追加検索キー候補リストを参照して、各追加検索キー候補の評価値を算出する追加検索キー選定手段と、前記追加検索キー選定手段により求めた評価値が高いものを追加検索キーとして出力する追加検索キー提示手段とを備えたものである。

【0010】第5の発明では、被検索テキストに対してインデックスを作成するインデックス作成行程と、前記被検索テキストから有効語候補を抽出し、各文書における出現頻度をカウントして有効語候補リストを作成する有効語候補抽出行程と、前記有効語候補抽出行程により抽出された有効語候補の出現頻度を文書別に保持する有効語候補頻度保持行程と、前記有効語候補リストを参照して、各有効語候補の評価値を算出し、評価値の高い有効語候補を有効語として、有効語リストを作成する有効語選定行程と、前記有効語リストを参照し、前記有効語候補頻度保持行程で保持された有効語候補の文書別頻度から有効語でない語を削除して、有効語の文書別頻度を作成する有効語頻度作成行程と、前記インデックスにより、検索文字列に一致する文字列を検索する検索行程と、前記有効語頻度保持行程で保持された有効語の文書

別頻度のうち、前記検索行程の検索結果の文書に相当するものから、追加検索キー候補リストを作成する追加検索キー候補リスト作成行程と、前記追加検索キー候補リストを参照して、各追加検索キー候補の評価値を算出する追加検索キー選定行程と、前記追加検索キー選定行程により求めた評価値が高いものを追加検索キーとして出力する追加検索キー提示行程とを実行するものである。

【0011】第6の発明では、被検索テキストを保持する被検索テキスト保持行程と、前記被検索テキスト保持行程で保持されている被検索テキストに対して、インデックスを作成するインデックス作成行程と、前記インデックス作成行程で作成したインデックスを保持するインデックス保持行程と、前記被検索テキスト保持行程で保持されている被検索テキストから有効語候補を抽出し、各文書における出現頻度をカウントして有効語候補リストを作成する有効語候補抽出行程と、前記有効語候補抽出行程により抽出された有効語候補の出現頻度を文書別に保持する有効語候補頻度保持行程と、前記有効語候補抽出行程により作成された有効語候補リストを保持する有効語候補リスト保持行程と、前記有効語候補リスト保持行程で保持された有効語候補リストを参照して、各有効語候補の評価値を算出し、評価値の高い有効語候補を有効語として有効語リストを作成する有効語選定行程と、前記有効語選定行程により作成された有効語リストを保持する有効語リスト保持行程と、前記有効語リスト保持行程で保持された有効語リストを参照し、上記有効語候補頻度保持行程で保持された有効語候補の文書別頻度から有効語でない語を削除して、有効語の文書別頻度を作成する有効語頻度作成行程と、前記有効語頻度作成行程により作成された有効語の文書別頻度を保持する有効語頻度保持行程と、検索を行なう文字列を保持する検索キー保持行程と、前記インデックス保持行程で保持されているインデックスを用いて、前記検索キー保持行程で保持されている検索文字列に一致する文字列を検索する検索行程と、前記検索行程により検索された検索結果を保持する検索結果保持行程と、前記有効語頻度保持行程で保持された有効語の文書別頻度のうち、前記検索結果保持行程で保持された検索結果の文書に相当するものから、追加検索キー候補リストを作成する追加検索キー候補リスト作成行程と、前記追加検索キー候補リスト作成行程により作成された追加検索キー候補リストを保持する追加検索キー候補リスト保持行程と、前記追加検索キー候補リスト保持行程で保持された追加検索キー候補リストを参照して、各追加検索キー候補の評価値を算出する追加検索キー選定行程と、前記追加検索キー選定行程により求めた評価値が高いものを追加検索キーとして保持する追加検索キー保持行程と、追加検索キー保持行程で保持された追加検索キーを出力する追加検索キー提示行程とを実行するものである。

【0012】第7の発明では、テキストから有効語候補

を抽出し、各文書における出現頻度をカウントして有効語候補リストを作成する有効語候補抽出行程と、前記有効候補抽出行程により抽出された有効語候補の出現頻度を文書別に保持する有効語候補頻度保持行程と、前記有効語候補リストを参照して、各有効語候補の評価値を算出し、評価値の高い有効語候補を有効語として、有効語リストを作成する有効語選定行程と、前記有効語リストを参照し、前記有効語候補頻度保持行程で保持された有効語候補の文書別頻度から有効語でない語を削除して、有効語の文書別頻度を作成する有効語頻度作成行程とを

【0013】第8の発明では、被検索テキストに対してインデックスを作成するインデックス作成行程と、前記インデックスにより、検索文字列に一致する文字列を検索する検索行程と、上記第7の発明の有効語頻度作成方法によって作成された有効語の文書別頻度のうち、前記検索行程の検索結果の文書に相当するものから、追加検索キー候補リストを作成する追加検索キー候補リスト作成行程と、前記追加検索キー候補リストを参照して、各追加検索キー候補の評価値を算出する追加検索キー選定行程と、前記追加検索キー選定行程により求めた評価値が高いものを追加検索キーとして出力する追加検索キー提示行程とを実行するものである。

【0014】第9の発明は、被検索テキストに対してインデックスを作成するインデックス作成行程と、前記被検索テキストから有効語候補を抽出し、各文書における出現頻度をカウントして有効語候補リストを作成する有効語候補抽出行程と、前記有効語候補抽出行程により抽出された有効語候補の出現頻度を文書別に保持する有効語候補頻度保持行程と、前記有効語候補リストを参照して、各有効語候補の評価値を算出し、評価値の高い有効語候補を有効語として、有効語リストを作成する有効語選定行程と、前記有効語リストを参照し、前記有効語候補頻度保持行程で保持された有効語候補の文書別頻度から有効語でない語を削除して、有効語の文書別頻度を作成する有効語頻度作成行程と、前記インデックスにより、検索文字列に一致する文字列を検索する検索行程と、前記有効語頻度保持行程で保持された有効語の文書別頻度のうち、前記検索行程の検索結果の文書に相当するものから、追加検索キー候補リストを作成する追加検索キー候補リスト作成行程と、前記追加検索キー候補リストを参照して、各追加検索キー候補の評価値を算出する追加検索キー選定行程と、前記追加検索キー選定行程により求めた評価値が高いものを追加検索キーとして出力する追加検索キー提示行程とを含むプログラムを格納したものである。

【0015】第10の発明は、被検索テキストを保持する被検索テキスト保持行程と、前記被検索テキスト保持行程で保持されている被検索テキストに対して、インデックスを作成するインデックス作成行程と、前記インデ

ックス作成行程で作成したインデックスを保持するインデックス保持行程と、前記被検索テキスト保持行程で保持されている被検索テキストから有効語候補を抽出し、各文書における出現頻度をカウントして有効語候補リストを作成する有効語候補抽出行程と、前記有効語候補抽出行程により抽出された有効語候補の出現頻度を文書別に保持する有効語候補頻度保持行程と、前記有効語候補抽出行程により作成された有効語候補リストを保持する有効語候補リスト保持行程と、前記有効語候補リスト保持行程で保持された有効語候補リストを参照して、各有効語候補の評価値を算出し、評価値の高い有効語候補を有効語として有効語リストを作成する有効語選定行程と、前記有効語選定行程により作成された有効語リストを保持する有効語リスト保持行程と、前記有効語リスト保持行程で保持された有効語リストを参照し、上記有効語候補頻度保持行程で保持された有効語候補の文書別頻度から有効語でない語を削除して、有効語の文書別頻度を作成する有効語頻度作成行程と、前記有効語頻度作成行程により作成された有効語の文書別頻度を保持する有効語頻度保持行程と、検索を行なう文字列を保持する検索キー保持行程と、前記インデックス保持行程で保持されているインデックスを用いて、前記検索キー保持行程で保持されている検索文字列に一致する文字列を検索する検索行程と、前記検索行程により検索された検索結果を保持する検索結果保持行程と、前記有効語頻度保持行程で保持された有効語の文書別頻度のうち、前記検索結果保持行程で保持された検索結果の文書に相当するものから、追加検索キー候補リストを作成する追加検索キー候補リスト作成行程と、前記追加検索キー候補リスト作成行程により作成された追加検索キー候補リストを保持する追加検索キー候補リスト保持行程と、前記追加検索キー候補リスト保持行程で保持された追加検索キー候補リストを参照して、各追加検索キー候補の評価値を算出する追加検索キー選定行程と、前記追加検索キー選定行程により求めた評価値が高いものを追加検索キーとして保持する追加検索キー保持行程と、追加検索キー保持行程で保持された追加検索キーを出力する追加検索キー提示行程とを含むプログラムを格納したものである。

【0016】第11の発明は、テキストから有効語候補を抽出し、各文書における出現頻度をカウントして有効語候補リストを作成する有効語候補抽出行程と、前記有効候補抽出行程により抽出された有効語候補の出現頻度を文書別に保持する有効語候補頻度保持行程と、前記有効語候補リストを参照して、各有効語候補の評価値を算出し、評価値の高い有効語候補を有効語として、有効語リストを作成する有効語選定行程と、前記有効語リストを参照し、前記有効語候補頻度保持行程で保持された有効語候補の文書別頻度から有効語でない語を削除して、有効語の文書別頻度を作成する有効語頻度作成行程とを含むプログラムを格納したものである。



【0017】第12の発明は、被検索テキストに対してインデックスを作成するインデックス作成行程と、前記インデックスにより、検索文字列に一致する文字列を検索する検索行程と、上記第11の発明の記録媒体中の有効語頻度作成行程によって作成された有効語の文書別頻度のうち、前記検索行程の検索結果の文書に相当するものから、追加検索キー候補リストを作成する追加検索キー候補リスト作成行程と、前記追加検索キー候補リストを参照して、各追加検索キー候補の評価値を算出する追加検索キー選定行程と、前記追加検索キー選定行程により求めた評価値が高いものを追加検索キーとして出力する追加検索キー提示行程とを含むプログラムを格納したものである。

【0018】

【発明の実施の形態】以下、図面を参照して本発明の実施の形態を説明する。

【0019】〔第1実施形態〕図1は、本発明の第1実施形態のテキスト検索装置の機能構成を示すブロック図である。

【0020】同図において、101は、被検索テキストを保持する被検索テキスト保持部である。102は、被検索テキスト保持部101に保持されている被検索テキストに対して、検索を高速に行なうためのインデックスを作成するインデックス作成部である。103は、インデックス作成部102で作成したインデックスを保持するインデックス保持部である。104は、被検索テキスト保持部101に保持されている被検索テキストから有効語候補を抽出し、各文書における出現頻度をカウントして有効語候補リストを作成する有効語候補抽出部である。有効語候補抽出部104は、被検索テキストを形態素解析し、その結果、名詞と解析された文字列を有効語候補とする。105は、有効語候補抽出部104により抽出された有効語候補の出現頻度を文書別に保持する有効語候補頻度保持部である。106は、有効語候補抽出部104により作成された有効語候補リストを保持する有効語候補リスト保持部である。有効語候補リストは、有効語候補ごとに各文書における出現頻度を保持している。有効語候補リストの例を図2に示す。

【0021】107は有効語選定部であり、有効語候補リスト保持部106に保持された有効語候補リストを参照して各有効語候補の評価値を算出し、評価値の高い有効語候補を有効語として有効語リストを作成する。有効語リストの例を図3に示す。評価値は、各有効語候補の重要度を示す値である。各文書の特徴付ける語が重要なので、検索対象文書内における各語の局在度を評価値に反映させるべきであるが、局在度そのものを評価値にすると、ある文書にたまたま一度出現した語の評価値が高くなってしまふ。そこで、ある程度の出現頻度があり、かつ局在度の高い語を有効語とするために、出現頻度と局在度の積を評価値とする。

【0022】有効語候補 $W_i$ を含む文書数を $M$ とすると、 $W_i$ の評価値 $E(W_i)$ は次の式により求められる。

【0023】

【数1】

$$E(W_i) = M \cdot \left(1 + \sum_{j=1}^M P_{i,j} \log_M P_{i,j}\right)$$

$$P_{i,j} = \frac{\text{文書 } D_j \text{ における } W_i \text{ の出現頻度}}{\text{文書 } D_j \text{ 中の有効語候補数}}$$

108は、有効語選定部107により作成された有効語リストを保持する有効語リスト保持部である。109は有効語頻度作成部であり、有効語リスト保持部108に保持された有効語リストを参照し、有効語候補頻度保持部105に保持された有効語候補の文書別頻度から有効語でない語を削除して、有効語の文書別頻度を作成する。

【0024】110は、有効語頻度作成部109により作成された有効語の文書別頻度を保持する有効語頻度保持部である。111は、検索を行なう文字列を保持する検索キー保持部である。112は検索部であり、インデックス保持部103に保持されているインデックスを用いて、検索キー保持部111に保持されている検索文字列に一致する文字列を検索する。113は、検索部112により検索された検索結果を保持する検索結果保持部である。

【0025】114は追加検索キー候補リスト作成部であり、有効語頻度保持部110に保持された有効語の文書別頻度のうち、検索結果保持部113に保持された検索結果の文書に相当するものから、追加検索キー候補リストを作成する。追加検索キー候補リストは、有効語候補リストと同様の構造で表現できる。115は、追加検索キー候補リスト作成部114により作成された追加検索キー候補リストを保持する追加検索キー候補リスト保持部である。

【0026】116は追加検索キー選定部であり、追加検索キー候補リスト保持部115に保持された追加検索キー候補リストを参照して、各追加検索キー候補の評価値を算出する。評価値は、各追加検索キー候補の再検索における有効度を示す値である。検索結果を絞り込むために再検索を行なうには、検索結果文書において局在度の高い語が有用と考えられるので、追加検索キー候補 $W$ を含む検索結果文書数を $M$ としたとき、 $W_i$ の評価値 $E(W_i)$ は次の式により求められる。

【0027】

【数2】

$$E(W_i) = M \cdot \left(1 + \sum_{j=1}^M P_{i,j} \log_M P_{i,j}\right)$$

$$P_{i,j} = \frac{\text{文書 } D_j \text{ 中における } W_i \text{ の出現頻度}}{\text{文書 } D_j \text{ 中の追加検索キー候補数}}$$

117は、追加検索キー選定部116により求めた評価値が高いものを追加検索キーとして保持する追加検索キー保持部である。

【0028】図4は、本実施形態に係るテキスト検索装置のハードウェア構成を示す図である。

【0029】同図において、201は、後述する制御手順を実現するプログラムを保持するROMである。202はRAMであり、被検索テキスト保持部101、有効語候補頻度保持部105、有効語候補リスト保持部106、有効語リスト保持部108、検索キー保持部111、検索結果保持部113、追加検索キー候補リスト保持部115、及び追加検索キー保持部117と、上記プログラムの動作に必要な記憶領域とを提供する。

【0030】203は、ROM201に保持されているプログラムに従って処理を行なう中央処理装置である。204はディスク装置であり、インデックス保持部103、有効語頻度保持部110を実現する。205はバスであり、上記の各構成を接続し、各構成間におけるデータの授受を可能とする。

【0031】次に、本装置の動作を説明する。

【0032】本実施形態の処理は、インデックス作成フェーズと検索フェーズとに大きく分かれる。以下、図5のフローチャートを参照して、本実施形態のテキスト検索装置におけるインデックス作成フェーズの手順を説明する。

【0033】ステップS301では、カウンタ1の初期化を行なう。カウンタ1は、処理の対象となっている被検索テキストの文書番号を示すもので、これを0に初期化する。そしてステップS302に移る。ステップS302では、カウンタ1が被検索テキストの文書数に達したか否かを判定して、達した場合はステップS308に移る。達していない場合は、ステップS303に移る。

【0034】ステップS303では、文書番号iの被検索テキストのインデックスを作成する。そしてステップS304に移る。ステップS304では、文書番号iの被検索テキストから、有効語候補となる文字列とその文書内出現頻度を抽出する。そしてステップS305に移る。ステップS305では、文書番号iの被検索テキストに出現する有効語候補とその頻度を保存する。そしてステップS306に移る。ステップS306では、文書番号iの被検索テキストの有効語候補に関する情報により、有効語候補リストを更新する。そしてステップS307に移る。

【0035】ステップS307では、カウンタ1の値を1増やす。そしてステップS302に戻る。ステップS308では、各有効語候補の評価値を算出する。そしてステップS309に移る。

【0036】ステップS309では、ステップS308で求めた評価値の高いもの、例えば、評価値が所定値以上のものを有効語とし、有効語リストを作成する。そしてステップS310に移る。ステップS310では、ステップS305で文書ごとに保存した有効語候補とその頻度から、有効語でないものを削除し、有効語とその頻度として保存する。そしてインデックス作成フェーズを終了する。

【0037】次に、図6のフローチャートを参照して、本実施形態における検索フェーズの概要を説明する。

【0038】ステップS401では、検索キーを含む文書を検索する。そしてステップS402に移る。ステップS402では、カウンタ1の初期化を行なう。カウンタ1は、ステップS401で得られた検索結果の結果番号を示すもので、これを0に初期化する。そしてステップS403に移る。ステップS403では、カウンタ1が検索結果数に達したか否かを判定して、達した場合はステップS406に移る。達していない場合はステップS404に移る。

【0039】ステップS404では、i番目の検索結果文書の有効語とその出現頻度を追加検索キー候補リストに追加する。そしてステップS405に移る。ステップS405では、カウンタ1の値を1増やす。そしてステップS403に戻る。ステップS406では、各追加検索キー候補の評価値を算出する。そしてステップS407に移る。ステップS407では、ステップS406で求めた評価値の高いものを追加検索キーとする。そして検索フェーズを終了する。

【0040】以上に述べたように、検索対象文書から追加検索キーの候補となる有効語を抽出し、検索結果文書における有効語の出現状況を反映した追加検索キーを提示することで、検索対象文書と検索結果文書に適合した再検索が容易に行なえるテキスト検索装置を実現することができる。

【0041】〔第2実施形態〕上記第1実施形態においては、インデックス作成フェーズと検索フェーズを同一装置で行なう場合について説明したが、これに限定されるものではない。例えば、インデックス作成フェーズだけを行なう装置であってもよい。図7はこの場合を示す本発明の第2実施形態の基本的な機能構成を示すブロック図である。

【0042】図7において、701は、被検索テキストを保持する被検索テキスト保持部である。702は、被検索テキスト保持部701に保持されている被検索テキストに対して、検索を高速に行なうためのインデックスを作成するインデックス作成部である。703は、イン

デックス作成部702で作成したインデックスを保持するインデックス保持部である。704は有効語候補抽出部であり、被検索テキスト保持部701に保持されている被検索テキストから有効語候補を抽出し、各文書における出現頻度をカウントして有効語候補リストを作成する。705は、有効語候補頻度保持部であり、有効語候補抽出部704により抽出された有効語候補の出現頻度を文書別に保持する。

【0043】706は有効語候補リスト保持部であり、有効語候補抽出部704により作成された有効語候補リストを保持する。707は有効語選定部であり、有効語候補リスト保持部706に保持された有効語候補リストを参照して、各有効語候補の評価値を算出し、評価値の高い有効語候補を有効語として、有効語リストを作成する。

【0044】708は有効語リスト保持部であり、有効語選定部707により作成された有効語リストを保持する。709は有効語頻度作成部であり、有効語リスト保持部708に保持された有効語リストを参照し、有効語候補頻度保持部705に保持された有効語候補の文書別頻度から有効語でない語を削除して、有効語の文書別頻度を作成する。710は有効語頻度保持部であり、有効語頻度作成部709により作成された有効語の文書別頻度を保持する。

【0045】インデックス保持部703に保持されたインデックスと有効語頻度保持部710に保持された有効語の文書別頻度は、通信回線を通したり、可搬記録媒体によって他の装置に移されて検索が行なわれる。

【0046】〔第3実施形態〕上述のようにして提供されたインデックスと有効語の文書別頻度を用いて検索を行なう装置であってもよい。図8はこの場合を示す本発明の第3実施形態の基本的な機能構成を示すブロック図である。

【0047】図8において、801は、図7に示す装置で作成されたインデックスを保持するインデックス保持部である。802は、検索を行なう文字列を保持する検索キー保持部である。803は検索部であり、インデックス保持部801に保持されているインデックスを用いて、検索キー保持部802に保持されている検索文字列に一致する文字列を検索する。

【0048】804は、検索部803により検索された検索結果を保持する検索結果保持部である。805は、図7に示す装置で作成された有効語の文書別頻度を保持する有効語頻度保持部である。806は追加検索キー候補リスト作成部であり、有効語頻度保持部805に保持された有効語の文書別頻度のうち、検索結果保持部804に保持された検索結果の文書に相当するものから、追加検索キー候補リストを作成する。

【0049】807は、追加検索キー候補リスト作成部806により作成された追加検索キー候補リストを保持

する追加検索キー候補リスト保持部である。808は追加検索キー選定部であり、追加検索キー候補リスト保持部807に保持された追加検索キー候補リストを参照して、各追加検索キー候補の評価値を算出する。809は追加検索キー保持部であり、追加検索キー選定部808により求めた評価値が高いものを追加検索キーとして保持する。

【0050】インデックス保持部801と有効語頻度保持部805には、上記第1実施形態の手順で作成されたインデックスと有効語頻度が、通信回線或いは可搬記録媒体等を介して提供され、保持される。

【0051】なお、本発明は図示の実施形態に限定されず種々の変形が可能である。例えば、その変形例としては次のようなものがある。

【0052】(1)上記第1実施形態では、検索結果を絞り込むために再検索を行なうものとして、追加検索キー候補の評価値を求めるよう説明したが、これに限定されるものでなく、以下の式により評価値を求め、検索条件を緩めて検索結果を増やすための再検索を行なう場合の追加検索キーを提示してもよい。

【0053】

【数3】

$$E(W_i) = M \cdot \left( - \sum_{j=1}^M P_{i,j} \log_M P_{i,j} \right)$$

$$P_{i,j} = \frac{\text{文書 } D_j \text{ 中における } W_i \text{ の出現頻度}}{\text{文書 } D_j \text{ 中の追加検索キー候補数}}$$

これは、検索結果文書に満遍なく出現する語は検索キーとの関連が深く、関連する語を検索条件に加えることで関連する文書を検索できるという考えに基づいている。

【0054】(2)ユーザが再検索意図を指定することにより、上記第1実施形態と上記変形例(1)の評価値計算法を切り替えることもできる。

【0055】(3)上記第1実施形態と上記変形例(1)の両方の評価値計算法による追加検索キーを再検索意図別に提示することもできる。

【0056】(4)上記第1実施形態では、ある程度の頻度で出現する語を有効語や追加検索キーにするために、評価値に出現頻度を掛けているが、出現頻度を指数とした単調増加関数の値を掛けても良い。

【0057】(5)上記第1実施形態では、各語の出現確率のエントロピーを1から減じた値を局在度としているが、これに限定されるものではなく、他の演算により局在度を求めてもよい。

【0058】(6)上記第1実施形態では、ステップS303でインデックス作成を行ない、その後に追加検索キーのための処理を行なうよう説明しているが、これに限定されるものではなく、ステップS306の後にイン

デックス作成を行なうようにしてもよい。

【0059】(7)上記第1実施形態では、必ず追加検索キーを提示するよう説明しているが、これに限定されるものではなく、追加検索キーを提示するかしないかをユーザが指定するようにしてもよい。

【0060】(8)上記第1実施形態では、形態素解析により有効語候補を抽出するよう説明しているが、これに限定されるものではなく、文字種の違いによって切り出した文字列を有効語候補としてもよい。

【0061】(9)上記第1実施形態では、名詞を有効語候補として抽出するよう説明しているが、これに限定されるものではなく、他の品詞の文字列や、複合語をも有効語候補として抽出してもよい。

【0062】(10)上記第1実施形態においては、被検索テキスト保持部101、有効語候補頻度保持部105、有効語候補リスト保持部106、有効語リスト保持部108、検索キー保持部111、検索結果保持部113、追加検索キー候補リスト保持部115、追加検索キー保持部117をRAMで、インデックス保持部103、有効語頻度保持部110をディスク装置で実現する場合について説明したが、これに限定されるものではなく、任意の記憶媒体を用いて実現してもよい。

【0063】(11)上記第1実施形態においては、各部を同一の計算機上で構成する場合について説明したが、これに限定されるものではなく、ネットワーク上に分散した計算機や処理装置などに分かれて各部を構成してもよい。

【0064】(12)上記第1実施形態においては、プログラムをROMに保持する場合について説明したが、これに限定されるものではなく、任意の記憶媒体を用いて実現してもよい。また、同様の動作をする回路で実現してもよい。

【0065】(13)本発明は、複数の機器から構成されるシステムに適用しても、1つの機器からなる装置に適用してもよい。前述した実施形態の機能を実現するソフトウェアのプログラムコードを記録した記録媒体を、システム或いは装置に供給し、そのシステム或いは装置のコンピュータ(またはCPUやMPU)が記録媒体に格納されたプログラムコードを読み出し実行することによっても、達成されることは言うまでもない。この場合、記録媒体から読み出されたプログラムコード自体が前述した実施形態の機能を実現することになり、そのプログラムコードを記録した記録媒体は本発明を構成することになる。

【0066】(14)プログラムコードを供給するための記録媒体としては、例えば、フロッピーディスク(登録商標)、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、CD-R、磁気テープ、不揮発性のメモ리카ード、ROMなどを用いることができる。

【0067】(15)コンピュータが読み出したプログ

ラムコードを実行することにより、前述した実施形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼働しているOSなどが実際の処理の一部または全部を行ない、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【0068】(16)記録媒体から読み出されたプログラムコードが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書き込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるCPUなどが実際の処理の一部または全部を行ない、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【0069】

【発明の効果】以上詳述したように、第1、第2及び第4の発明であるテキスト検索装置によれば、検索対象文書と検索結果文書に適合した再検索を容易に行なうことが可能になる。

【0070】第5、第6及び第8の発明であるテキスト検索方法によれば、上記第1、第2及び第4の発明と同等の効果を奏する。

【0071】第9、第10及び第12の発明であるコンピュータ読み取り可能な記録媒体によれば、コンピュータによりプログラムを読み出して実行することにより上記第1、第2及び第4の発明と同等の効果を奏する。

【0072】第3の発明である有効語頻度作成装置によれば、テキストにおける有効語の文書別頻度を簡易かつ的確に作成することが可能になる。

【0073】第7の発明である有効語頻度作成方法によれば、上記第3の発明と同等の効果を奏する。

【0074】第11の発明であるコンピュータ読み取り可能な記録媒体によれば、コンピュータによりプログラムを読み出して実行することにより上記第3の発明と同等の効果を奏する。

【図面の簡単な説明】

【図1】本発明の第1実施形態のテキスト検索装置の機能構成を示すブロック図である。

【図2】有効語候補リストの例を示す図である。

【図3】有効語リストの例を示す図である。

【図4】実施形態に係るテキスト検索装置のハードウェア構成を示す図である。

【図5】第1実施形態におけるインデックス作成フェーズの概要を示すフローチャートである。

【図6】第1実施形態における検索フェーズの概要を示すフローチャートである。

【図7】本発明の第2実施形態の基本的な機能構成を示すブロック図である。

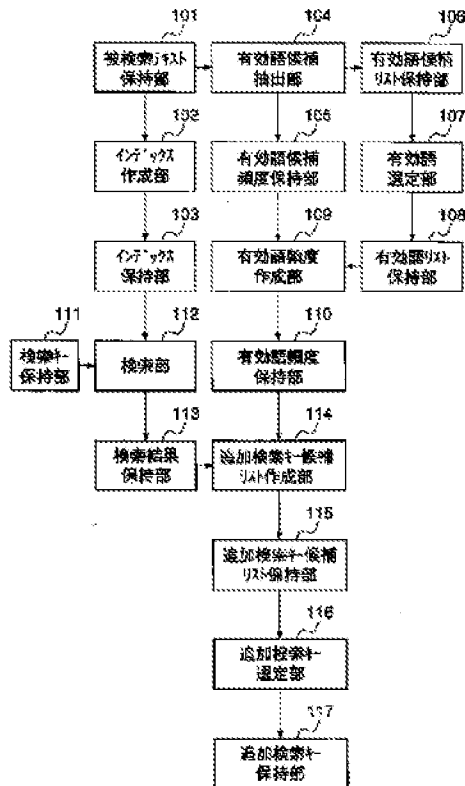
【図8】本発明の第3実施形態の基本的な機能構成を示すブロック図である。

## 【符号の説明】

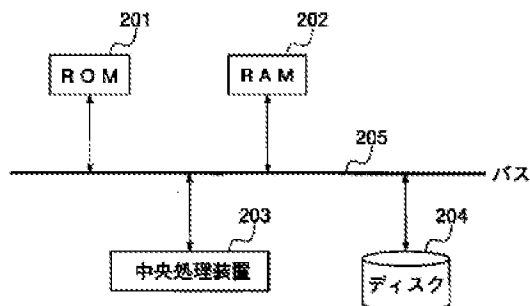
101 被検索テキスト保持部  
102 インデックス作成部  
103 インデックス保持部  
104 有効語候補抽出部  
105 有効語候補頻度保持部  
106 有効語候補リスト保持部  
107 有効語選定部  
108 有効語リスト保持部

109 有効語頻度作成部  
110 有効語頻度保持部  
111 検索キー保持部  
112 検索部  
113 検索結果保持部  
114 追加検索キー候補リスト作成部  
115 追加検索キー候補リスト保持部  
116 追加検索キー選定部  
117 追加検索キー保持部

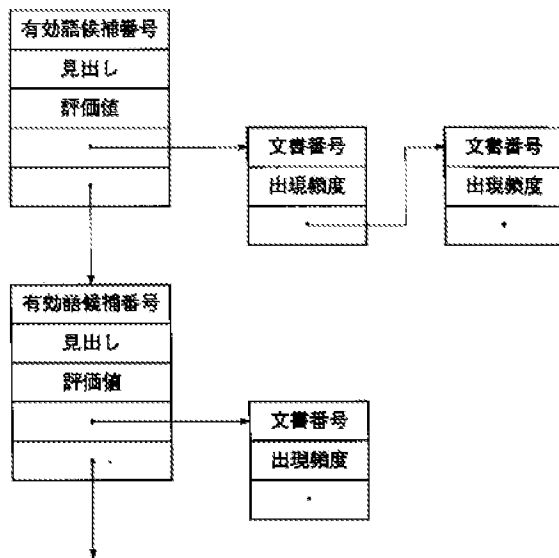
【図1】



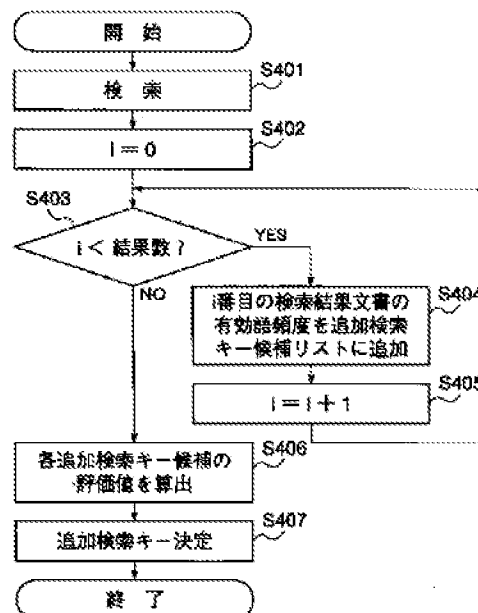
【図4】



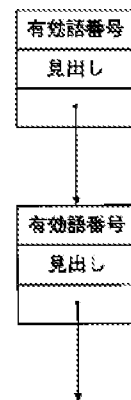
【図2】



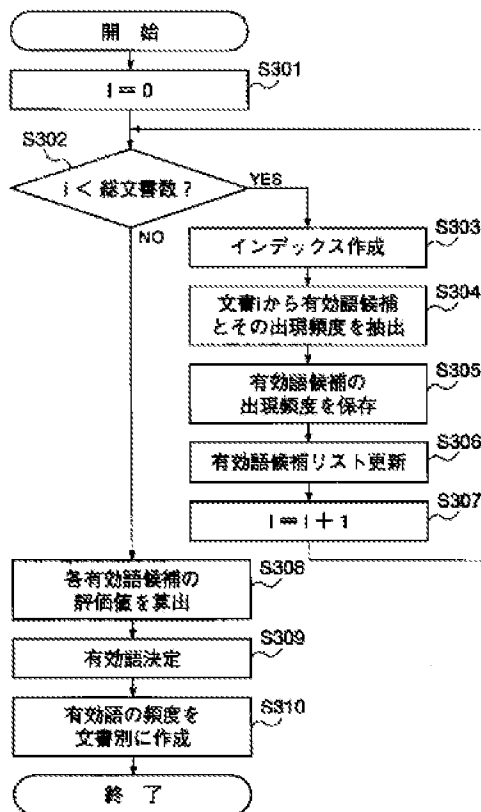
【図6】



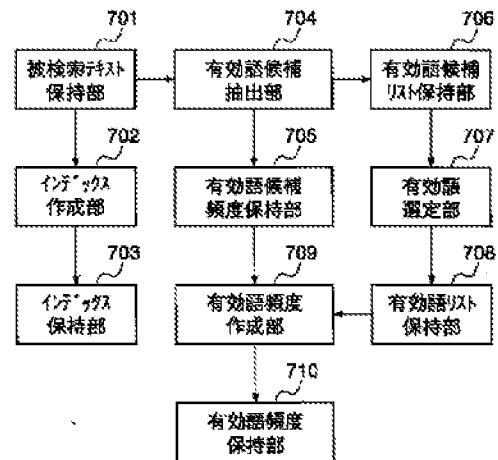
【図3】



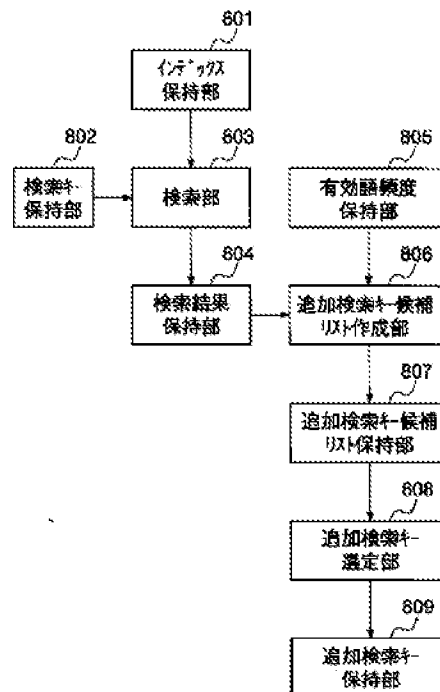
【図5】



【図7】



【図8】



フロントページの続き

(72)発明者 上田 隆也  
東京都大田区下丸子3丁目30番2号 キヤ  
ノン株式会社内

(72)発明者 池田 裕治  
東京都大田区下丸子3丁目30番2号 キヤ  
ノン株式会社内

Fターム(参考) 5B675 ND03 NK02 NK14 NK31 PP12  
PP22 PQ27 PQ35 PQ04